

DISPERSAL SPECIAL FEATURE

Measuring long-distance seed dispersal in complex natural environments: an evaluation and integration of classical and genetic methods

F. Andy Jones^{1,2*} and Helen C. Muller-Landau^{1,2}¹Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama; and ²Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN 55018, USA

Summary

1. Seed dispersal is a critical life stage of plants, yet accurate measurement of dispersal distances has been difficult in natural systems. Genetic techniques for matching dispersed seeds to maternal trees provide valuable data on dispersal events. Questions remain regarding how best to estimate the population seed dispersal distance distributions from such data and how these estimates compare with classical non-genetic estimates based on seed trap data alone.

2. Using simulated data, we compared seed shadow estimates obtained via standard inverse modelling of seed arrival into seed traps within mapped stands (summed seed shadow, SSS) with estimates from four models using genetic matches: direct fitting of the observed distribution of distances in the genotyped sample (observed distance, OBS), direct fitting inversely weighted by sampling intensity (OBSw), inverse modelling of numbers of seeds of each genotype in each trap (gene shadow model, GSM), and inverse modelling of frequencies of each genotype in each trap (competing sources model, CSM). We further explored how consideration of immigrant seed rain from unmapped and ungenotyped trees outside the stand affected dispersal estimates, denoting these variants SSSi, GSMi and CSMi.

3. We applied these models to an empirical data set for the Neotropical tree *Jacaranda copaia*, using a hierarchical Bayesian model to incorporate variation in fecundity among trees.

4. Fits to simulated data sets showed that OBS and SSS estimates were strongly biased, while SSSi, GSMi and CSMi were mildly biased. Root mean square errors varied as $OBS \gg SSS > OBSw > CSMi > GSMi > SSSi > CSM > GSM$.

5. Comparing results for *Jacaranda* under the three best models, mean posterior estimates of dispersal distances varied as $SSSi < GSM < CSM$, but credible intervals overlapped among all three models, demonstrating agreement that long-distance dispersal is common.

6. *Synthesis.* Here we reconcile two general approaches used to study seed dispersal distances in natural communities. Genetic and non-genetic approaches can both provide good estimates of seed dispersal provided that analyses of genetic data take account of any deviation from random selection of seeds for genotyping, and that SSS models consider immigrant seed rain in the analysis whenever there is reason to think its contribution is non-zero. The use of the models presented here should provide better estimates of population-level dispersal distance distributions in genetic and classical seed dispersal studies.

Key-words: competing sources model, dispersal kernel, dispersal limitation, hierarchical Bayes, microsatellite, *Jacaranda copaia*, long-distance seed dispersal, summed seed shadow model, tree fecundity

*Correspondence author. E-mail: jonesf@si.edu

Introduction

Direct measures of seed dispersal distances and plant fecundity are difficult to obtain in natural systems, especially in closed canopy forests where seed shadows overlap (Bullock *et al.* 2006). Recently, a novel genetic technique that uses genotypes of maternal tissue attached to seeds (endocarps, fruit, or seed wings) to unequivocally match dispersed seeds to maternal source trees has been used to measure seed dispersal distances within large complex natural systems (Godoy & Jordano 2001; Grivet *et al.* 2005; Jones *et al.* 2005; Pairon *et al.* 2006; Hansen *et al.* 2007; Isagi *et al.* 2007; Jordano *et al.* 2007). Although the method is highly accurate and can be used to measure seed movement across large spatial scales, the technical difficulty of obtaining such estimates means that its widespread application may be limited. The complexity of the issues surrounding fitting seed dispersal kernels to genetic data has only recently been appreciated (Robledo-Arnuncio & García 2007). Finally, classical non-genetic estimates of dispersal kernels have become common in studies of dispersal and recruitment limitation (Ribbens *et al.* 1994; Clark *et al.* 1999; Muller-Landau *et al.* this issue; Pouden *et al.* this issue; Schurr *et al.* this issue). What do dispersal kernels fit with the aid of genetic identification of seed parentage tell us that classical estimates do not? How well do these models estimate the frequency and intensity of long-distance dispersal (LDD) and immigration within natural populations? Knowledge of the frequency of immigration by seed may be of particular importance for studies concerned with the conservation of sensitive species (Grivet *et al.* 2005), species ability to disperse among fragmented habitats (Hansen *et al.* 2007), species response to altered or absent mutualist disperser assemblages (Wang *et al.* 2007), and forecasting species response to climatic change.

The nature of genetic seed dispersal data poses several challenges in fitting dispersal kernels. The distribution of observed dispersal distances from source plant to seed trap in a particular sample of genotyped seeds is not necessarily a good estimate of the full dispersal kernel, for several reasons (Robledo-Arnuncio & García 2007). First, genetic methods often show that some seeds are not the offspring of any of the genotyped mothers and are immigrants. These censored data provide a minimum dispersal distance estimates (distance to plot edge, for example) but actual distances remain unknown. Second, parental multilocus genotypes may not be unique, so that two or more possible mothers are identified for some individuals. In both these cases, the common use of the minimum possible dispersal distance leads to a downward bias in the dispersal kernel (Jones *et al.* 2005). Third, sampling schemes developed to specifically maximize information on LDD or local dispersal (for example, placing seed traps with respect to distance to source trees, or genotyping a lower proportion of seeds falling near source trees) will bias the distribution of genotyped seeds away from that found in a random sample of the full population (Hardesty *et al.* 2006). Fourth, even if seed traps are randomly placed with respect to source trees and all seeds falling into the traps are genotyped and

successfully matched to a single parent, the distribution of dispersal distances of seeds sampled from seed traps depends on the distribution of trees with respect to those seed traps, tree fecundities, and the dispersal kernel (Robledo-Arnuncio & García 2007). These estimates can potentially be quite different (see box 2 of Nathan & Muller-Landau 2000). Uncertainty in actual source trees (1&2) and systematic or random deviations introduced by the sampling strategy (3 and 4) can be addressed through maximum likelihood and hierarchical Bayesian models to arrive at unbiased estimates of the distribution of dispersal distances in the population.

One advantage of high-resolution genetic methods is that they allow one to identify immigrant seeds arriving from outside the area of mapped and genotyped trees. Standard inverse models that estimate seed shadows from seed rain within mapped stands (summed seed shadow, SSS method) often assume that immigrant seed rain is zero (Ribbens *et al.* 1994; Clark *et al.* 1998; Clark *et al.* 2004). However, such an assumption is not necessary; it is possible to numerically integrate the dispersal kernel over the area outside the mapped plot to calculate immigrant seed rain under a candidate parameter set (SSSi method; Muller-Landau *et al.* 2002; Dalling *et al.* 2002; Muller-Landau *et al.* this issue). Similarly, estimates of dispersal kernels from genetic data can use the information on immigrant seeds to exclude them from the analysis (Robledo-Arnuncio & García 2007), or fit the numbers of immigrant seeds with non-immigrant seeds through integration of seed rain expected from outside the mapped area. To date, no studies have compared models that did and did not include immigrant seed rain for either genetic analyses or SSS estimates.

Here, we apply models that build upon previous empirical and theoretical work to fit dispersal kernels to dispersal distances determined from seed trap data and genotyped seeds across scales up to 1 km. Using simulated data sets and a case study of the Neotropical tree *Jaccaranda copaia* (Aubl.) D. Don, we compare the classical model of fitting dispersal kernels to densities of seeds falling within traps, the SSS model (Ribbens *et al.* 1994; Clark *et al.* 1998, 2004), to several models developed for fitting dispersal kernels to dispersal distances determined through genetic identification.

First, we use dispersal kernel estimates that are fit directly to observed distances (the OBS model, Jones *et al.* 2005), and a variant in which the OBSs are weighted inversely to sampling intensity (OBSw). Next, we apply the SSS inverse modelling framework to genetic data by summing expected seeds across genotypes and traps, thereby fitting the densities of genotypes within each trap (gene shadow model, GSM). Finally, we fit models based upon the frequencies of individual genotypes in each seed trap using a multinomial error distribution, the competing sources model (CSM, Robledo-Arnuncio & García 2007). For each case, we compare models that incorporate estimates and/or data on immigrant seed rain with those that do not, distinguishing the immigrant-inclusive models with the suffix 'i' (e.g. SSS vs. SSSi). We quantify the bias and error of each model as fitted to the simulated data sets, thereby demonstrating that the OBS and SSS models are strongly biased, and that these biases can be

greatly reduced by appropriate weighting (OBSw) and by incorporating immigrant seed rain (SSSi), respectively. We further show that the best estimates are obtained by combining genetic data with inverse models through the GSM or CSM. We then compare empirical estimates obtained by applying the three best models (SSSi, GSM and CSM) to our case study, and find them to be broadly congruent.

Methods

EMPIRICAL DATA SET: STUDY SITE AND SPECIES

This study took place in closed-canopy, mature, moist tropical forest on Barro Colorado Island, Panama (Leigh *et al.* 1996), and specifically on the 50-ha forest dynamics plot (FDP) located on the centre of the island (Hubbell & Foster 1983). All trees > 1 cm in diameter on this plot have been mapped, measured, and identified to species every 5 years since 1980. For this study, we also mapped reproductive-sized individuals of our study species within 120 m of the borders of the plot, for a total sampled area of 91.76 ha. Since 1987, 200 seed traps, each 0.5 m² in area, located along the trails of this plot have been censused weekly with all seeds and fruits counted and identified to species. In 2002, because tree fall gaps were not well represented in the original design, we placed an additional 98 traps within over 50 new tree fall gaps of varying size across the FDP. We use seed rain into all 298 traps in 2002 and genetic data from subsamples of seeds from the same year for the analyses here.

Jacaranda copaia (Bignoniaceae) is a large emergent canopy tree (up to 45 m tall) characteristic of neotropical moist forests and ranges from Belize to Brazil and Bolivia (Croat 1978). It is a light demanding colonizer of large tree fall gaps (Brokaw 1985). Its small seeds (4.7 mg) are dispersed by wind in August and September. There were 205 reproductive-sized *J. copaia* within the BCI FDP and an additional 99 reproductive-sized individuals within 120 m of the FDP in 2002. Microsatellite genetic markers were developed for *J. copaia* in order to study seed movement (Jones & Hubbell 2003). All adult trees and a subset of the seeds falling into the traps were genotyped at four microsatellite markers (Jones *et al.* 2005; Jones & Hubbell 2006). In particular, up to five seeds (fewer if fewer seeds arrived) were sampled from every trap every week during the dispersal season. A total of 445 seeds were successfully genotyped of 5165 seeds that were captured in traps in 2002 (Fig. 1a,b). A summary of tree, trap, seed, and genotype information is found in Table 1.

SIMULATED DATA SETS

To assess the bias and error of the different fitting methods against a known standard, we simulated 100 replicate data sets with known dispersal and fecundity parameters and seed source identities. We used data set size and parameter values similar to those we obtained for *J. copaia*. Specifically, we used the true locations and sizes of *J. copaia* trees on the plot (1240 × 740 m), and simulated additional source trees in a larger area extending to 2000 m from the plot edges, for a total area of discrete source trees encompassing 5240 × 4740 m. The added trees were independently randomly placed at a density equal to that of *J. copaia* trees on the plot, and their sizes were drawn with replacement from the distribution of sizes of *J. copaia* trees on the plot. We simulated seed rain from all these source trees to 300 seed traps randomly placed within the plot, assuming a fecundity of one seed produced per cm² basal area (the same for all trees), dispersal according to a two-dimensional Student's *t* distribution (Clark *et al.* 1999)

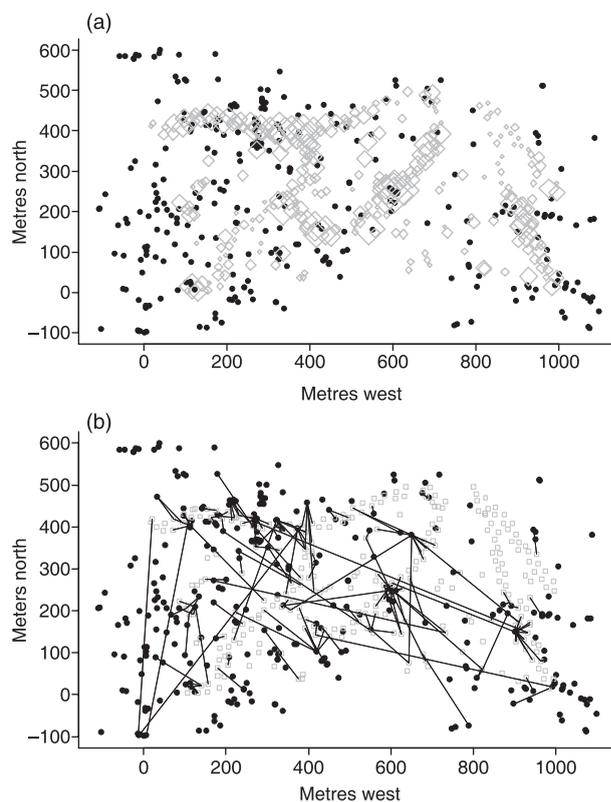


Fig. 1. (a) Map of *Jacaranda* seed rain captured in seed traps on the Forest Dynamics plot on Barro Colorado Island, Panama. Grey squares are locations of seed traps ($n = 298$) that are scaled to represent the log (seed number + 1). Dark circles are location of reproductive sized *Jacaranda copaia* ($n = 305$). (b) Solid lines represent the movement of seeds from trees to seed traps as measured using genetic markers.

Table 1. Summary of study area, organisms, and genetic data used in the empirical analysis of dispersal kernel fits. Detailed descriptions of study sites, organisms, and methods can be found in Jones *et al.* (2005)

Variable	<i>Jacaranda copaia</i>
Extent of mapped plot (ha)	191.76
Mechanism of dispersal	Wind
Total adult trees	304
Mean density ha ⁻¹	3.6
Minimum reproductive size (cm diameter)	20
Number of seed traps	298
Total seeds captured	5165
Mean number of seeds trap ⁻¹ (SD)	17.32 (46.12)
Number of seeds genotyped (% of total)	445 (8.6)
Number of microsatellite loci	4
Average alleles locus ⁻¹	11.7
Average expected heterozygosity (H_e)	0.81
Probability of Identity	8.13×10^{-6}

$$F(r) = \frac{p}{\pi u \left(1 + \frac{r^2}{u}\right)^{p+1}} \quad \text{eqn 1}$$

with $p = 100$ and $u = 0.5$, and a negative binomial distribution of actual seeds arriving from a given tree into a given trap around the

expected distribution, with dispersion parameter $k = 100$. The locations of the seed traps as well as the locations and sizes of the source trees off the plot were chosen separately for each simulation. This produced an average of 7953 seeds arriving to seed traps in total. A maximum of 20 randomly chosen seeds arriving to each trap were 'genotyped' to generate genetic data. The data sets to which the models were fit consisted of the true source trees on the plot, the locations and numbers of seeds arriving into the seed traps on the plot, and/or the genetic data on a subset of the seeds arriving into seed traps.

For the simulated parameter values, 0.5% of seeds are expected to go beyond 2000 m, and thus to arrive from beyond 2000 m in a landscape of continuous source trees. We therefore not only simulated seed rain from the discretely placed source trees extending 2000 m beyond the plot edge, we also numerically integrated expected seed rain from outside the area of discrete source trees.

GENERAL MODEL STRUCTURE

Classical inverse modelling simultaneously estimates fecundity and dispersal parameters without the direct observation of individual tree seed production and measurement of the distances that individual seeds move (Ribbens *et al.* 1994; Clark *et al.* 1998; Clark *et al.* 1999; Dalling *et al.* 2002; Clark *et al.* 2004). These models use tree locations, tree diameters, seed trap locations, and seed trap counts to parameterize spatially explicit models of seed dispersal. The models estimate annual seed deposition to any location within the study area as a function of a structural dispersal kernel scaled to the fecundity of individual trees. Following Clark *et al.* (2004), we incorporate a hierarchical Bayesian model in our analyses to account for individual variation in seed production (fecundity) found in forest trees.

We performed the modelling using programs run in the R operating environment <www.cran.r-project.org>. All programs used in the analyses are available from HCM upon request. All of the methods we compare fitted the dispersal kernel $F(r)$ defined here as the probability of a seed arriving per unit area at distance x (also known as the two-dimensional dispersal kernel). Models also simultaneously estimated parameters related to fecundity and clumping of seed deposition. We report results for fits of the two-dimensional Student's t distribution (Clark *et al.* 1999) to the dispersal kernel (eqn 1). Preliminary tests showed that this distribution fit the data for our case study species significantly better than any of the tested alternatives (Weibull, lognormal, Gaussian or exponential). To fit the models, we used a reparameterization of this model in which the fitted parameters were log transforms of the original parameters, because fits converged better under this parameterization. We then back-transformed these fitted parameters into the original form for the reporting of results. To interpret the parameter values, we also report estimated median dispersal distances, calculated from the parameters above as $\sqrt[3]{(u^2)^p - 1}$, and mean dispersal distances calculated for $P > 0.5$ as $\sqrt[3]{u\{\Gamma(1.5)\Gamma(p - 0.5)/[\Gamma(p + 1)]\}}$ (for $P < 0.5$, the mean dispersal distance is infinite). Note further that the total proportion of seeds going beyond distance r can be calculated as $[1 + (r^2/u)]^{-p}$, which is useful for calculating the quantity of immigrant seed rain.

Models incorporating variation in fecundity among individual trees were hierarchically structured, so that individual tree fecundity parameters β_i were fit along with hyperparameters for their mean β^* and SD σ , that is,

$$Q_i = \exp(\beta_i)b_i$$

$$\beta_i \approx \text{Gaussian}(\mu, \sigma)$$

Here Q_i is the fecundity of tree i in seeds produced per year, β_i is the fecundity parameter of tree i , and b_i is the basal area of tree i (in mm^2 basal area). Under this model, seed production per unit basal area varies lognormally among trees (because the log of the individual seed production per unit basal area is normally distributed). In the case of one fitting method (the CSM), we fit the SD of the fecundity parameter without fitting the mean (see explanation in following sections).

THE SSS: INVERSE MODELLING FROM NUMBERS OF SEEDS IN TRAPS

Under this model, we estimated dispersal and fecundity parameters from the location and number of seeds in each seed trap and the basal area and locations of adults alone, that is, without using any genetic information (Ribbens *et al.* 1994; Tufto *et al.* 1997; Clark *et al.* 1998). The expected seed rain into a trap is calculated as the sum of contributions from every conspecific adult tree on the plot, with each contribution determined by that tree's distance from the trap and its basal area. Thus, the expected seed rain, s_j , into a trap j , is

$$\hat{s}_j = x \sum_{\text{trees } i} Q_i F(r_{ij}) \quad \text{eqn 2}$$

where x is the area of the trap (to convert the total from seeds per unit area to seeds per trap). For the immigrant version of this model, SSSI, we further considered the potential input of seeds from source tree outside the plot, by numerically integrating over the area outside the plot, assuming that the density of source trees there was the same as found on the plot (Dalling *et al.* 2002; Muller-Landau *et al.* 2002; Muller-Landau *et al.* 2004):

$$\hat{s}_j = x \left[\sum_{\text{trees } i} Q_i F(r_{ij}) + \frac{\sum Q_i}{a} \iint_{\text{area off plot}} F\left(\sqrt{(x_i - x)^2 + (y_j - y)^2}\right) dx dy \right] \quad \text{eqn 3}$$

where a is the total mapped area, x_j and y_j are the coordinates of the trap, and x and y are the coordinates of points outside the plot. Note that the expected number of immigrant seeds varies among traps depending on their positions x_j and y_j ; traps nearer to the edge are expected to receive more trees than traps farther away. The likelihood of the data showing s_j seeds arriving in trap j for all traps combined is then

$$\prod_j \Pr(s_j | \hat{s}_j).$$

We used a negative binomial for the error distribution with a clumping parameter which we denote k (Hilborn and Mangel 1997; Clark *et al.* 1998).

THE OBS MODEL: FITTING DIRECTLY TO THE DISTRIBUTIONS OF DISPERSAL DISTANCES

An obvious model for estimating the dispersal kernel from genetic data linking dispersed seeds with parent plants is to fit the dispersal kernel directly to the distribution of observed dispersal distances in the sampled data (Jones *et al.* 2005; Hardesty *et al.* 2006). For seeds whose maternal genotypes match one parent tree, the distance from the seed sampling point to the parent was the dispersal distance. For seeds matching more than one parent tree, the minimum distance

was used (to be conservative). For seeds that match none of the parent trees (immigrants), the minimum distance to the edge of the area beyond which adult trees were not genotyped was used as the dispersal distance. A probability distribution was fit directly to the resulting set of distances, a model previously used by Jones *et al.* (2005).

THE GSM MODEL: INVERSE MODELLING FROM THE NUMBERS OF EACH GENOTYPE IN TRAPS

The inverse modelling approach applied to data on total numbers of seeds in traps can be extended to numbers of seeds of each genotype in a sample from each trap. Let G_i be the genotype of tree i . The expected number of seeds of genotype g to be identified among the seeds sampled from seed trap i , \hat{s}_{ig} , can then be calculated simply by summing the expected contributions of all parent trees j having genotype $G_j = g$:

$$\hat{s}_{gj} = c_j x \sum_i^{G_i=g} Q_i F(r_{ij}) \tag{eqn 4}$$

where c_j is the proportion of all seeds in trap j that are genotyped. Note that the implicit assumption of this equation is that all seeds with genotype g are offspring of one of the identified mothers having the matching genotype. Thus, it is assumed that none of these seeds could have come from non-genotyped parent trees. The total likelihood of all the genotyped data for seeds matching known trees alone is

$$\prod_j \prod_i \Pr(s_{ij} | \hat{s}_{ij})$$

where s_{ij} is the observed number of seeds of genotype i in trap j . As with the seed shadow model, we assumed a negative binomial error distribution.

Similarly, for the immigrant version of this model, GSMi, we can calculate the expected number of seeds whose genotype does not match that of any parent tree on the plot, \hat{s}_{*j} , under the assumption that all trees off the plot have non-matching genotypes, as

$$\hat{s}_{*j} = c_j x \frac{\sum Q_i}{a} \iint_{\text{area off plot}} F\left(\sqrt{(x_j - x)^2 + (y_j - y)^2}\right) dx dy \tag{eqn 5}$$

That is, we assume all the seeds coming from parents off the plot have genotypes that match none of the parent trees on the plot. The total likelihood of all the genotyped data is then

$$\prod_j \prod_i \Pr(s_{ij} | \hat{s}_{ij}) \prod_j \Pr(s_{*j} | \hat{s}_{*j})$$

where s_{*j} is the observed number of seeds in trap j whose genotype matches none of the parents on the plot.

THE CSM: INVERSE MODELLING FROM THE RELATIVE FREQUENCIES OF GENOTYPES IN TRAPS

The CSM introduced by Robledo-Arnuncio & García (2007) analyses the probabilities of observing different frequencies of genotypes in each trap rather than absolute numbers of each genotype in each trap. The likelihood of observed numbers of seeds of each genotype in a trap relative to these expected frequencies is evaluated via multinomial error distribution.

In the original version of this model, the observed and expected frequencies are calculated without consideration of immigrant seed

rain. That is, seeds whose genotype does not match any source tree in the study area are dropped from the analysis. Further, source trees are assumed to have the same fecundity. Extending this model to incorporate the possibility that multiple parent trees have the same genotype as above, and allowing parents to differ in fecundity, the expected frequency of seeds of genotype g among all genotyped seeds matching known source trees in trap j , \hat{p}_{gj} , becomes

$$\hat{p}_{gj} = \frac{\sum_i^{G_i=g} Q_i F(r_{ij})}{\sum_{\text{trees } k} Q_k F(r_{kj})} \tag{eqn 6}$$

The total likelihood of the data is then

$$\prod_j \Pr(p_{1j}, p_{2j}, \dots, p_{Gj} | \hat{p}_{1j}, \hat{p}_{2j}, \dots, \hat{p}_{Gj})$$

evaluated using a multinomial distribution.

To incorporate immigrant seed rain (CSMi), we first calculate the expected proportion of immigrant seeds, as before. We then evaluate the frequency of each matched genotype among all genotyped seeds in each trap (including presumed immigrant seeds of unmatched genotype), and the frequency of unmatched genotypes in each trap, against the expected values using the multinomial. Specifically, the expected frequency of the seeds of genotype g in trap j , \hat{p}_{gj} , becomes

$$\hat{p}_{gj} = \frac{\sum_i^{G_i=g} Q_i F(r_{ij})}{\sum_{\text{trees } i} Q_i F(r_{ij}) + \frac{\sum Q_i}{a} \iint_{\text{area off plot}} F\left(\sqrt{(x_j - x)^2 + (y_j - y)^2}\right) dx dy} \tag{eqn 7}$$

The expected frequency of seeds in trap j having a genotype that does not match any of the trees on the plot, \hat{p}_{*j} , is likewise

$$\hat{p}_{*j} = \frac{\frac{\sum Q_i}{a} \iint_{\text{area off plot}} F\left(\sqrt{(x_j - x)^2 + (y_j - y)^2}\right) dx dy}{\sum_{\text{trees } i} Q_i F(r_{ij}) + \frac{\sum Q_i}{a} \iint_{\text{area off plot}} F\left(\sqrt{(x_j - x)^2 + (y_j - y)^2}\right) dx dy} \tag{eqn 8}$$

These frequencies are then evaluated with multinomial error.

For many models of fecundity, one population-level fecundity parameter will cancel out of eqns 6, 7 and 8, thus reducing the number of parameters estimated. For example, if fecundity is equal to a constant times tree size as in our simulations, this constant will cancel out. For our model of fecundity incorporating variation among individuals, the hyperparameter β^* for the mean of the individual fecundity parameters effectively disappears and we thus fit only the SD σ of individual fecundity parameters.

FITTING THE MODELS TO SIMULATED DATA

To assess error and bias of each fitting method described above, with and without consideration of immigrant seed rain, we fitted each of the simulated data sets using each model. The fitted model assumed identical seed production per basal area across all source trees, the 2Dt dispersal kernel, and negative binomial variation in actual seed rain around expected values, in all cases as in the

simulations. We also fitted a 2Dt kernel in which the p parameter was set equal to 1, to test the recommendation that this is the better approach when the true value of p is below 1. Because posterior estimates of parameters for fits using the Gibbs sampler were essentially identical to point estimates obtained using local optimization, because we were interested only in the best-fit values for this exercise, and because the local optimization methods were many times faster, we report analyses of the results of the local optimization fits. We calculated the median, 95% confidence interval, bias (mean deviation from the true value), coefficient of variation, and relative root mean squared error (RRMSE, the root mean squared error divided by the true value) of each fitted parameter under each fitting method.

FITTING THE MODELS TO EMPIRICAL DATA

We used hierarchical Bayesian models to fit the models to the empirical data set (Gelman *et al.* 1995; Clark 2007). We used non-informative priors for all fitted parameter estimates. We used the medians of the posterior distributions of each parameter as our best estimate, and report 95% credible intervals. Depending on model time to convergence, we ran three independent Markov chains of at least 3000–10 000 iterations with different starting parameters. The total number of iterations and the burn-in period was determined separately for each models based upon the convergence dynamics. We used Gelman & Rubin's (1992) and Brooks & Gelman's (1998) potential scale reduction factor (PSRF), which measures the ratio of variance within and among chains, to assess model convergence. A model was assumed to have converged if the PSRF was < 1.1. Convergence calculations and visualizations were performed using the CODA package (version 0.11–2) in the R environment.

Results

The GSM and CSM both performed well on the simulated data sets, with insignificant bias and low relative root mean squared errors of parameter estimates (Table 2). The SSSi, GSMi and CSMi also performed well, but had small significant biases in one or both dispersal parameters, and larger errors in general. The OBSw model had only a small bias in the second dispersal parameter, but had much larger errors than the inverse modelling genetic estimates. The SSS and OBS models both had very large biases and very large errors. The SSS model in particular greatly overestimated fecundity and showed extreme error and bias. Furthermore, the SSS only converged 75% of the time, while the GSM converged 99% of the time and all other models converged 100% of the time. Fitting a 1-parameter 2Dt function with the p parameter fixed at 1 resulted in extreme overestimation of the u parameter (median estimates varied from 501 for the GSM to 4089 for the SSS), increases the confidence intervals on u by factors of 3.4 (CSM) to 1581 (SSS) and none of the confidence intervals overlapped the true value (100). Estimates of the fecundity parameter under the 1-parameter fits were not as bad as the dispersal estimates, but confidence intervals overlapped the true value only in the case of the SSSi, and biases, errors, and CI size were higher for all methods except SSS (data not shown).

The SSSi, GSM and CSM estimates of dispersal parameters for *Jacaranda* did not differ significantly, as their 95%

Table 2. Results of fits to simulated data sets for different kernel estimation models

Fit type	Prop	Fecund					u					p				
		Estimate (95% CI)	Bias	CV (%)	RRMSE	Estimate (95% CI)	Bias	CV (%)	RRMSE	Estimate (95% CI)	Bias	CV (%)	RRMSE			
OBS	1					566 (396.8, 813)	467.9***	17.61	4.784	0.665 (0.602, 0.732)	0.164***	5.11	0.335			
OBSw	1					124.8 (43.4, 224.7)	24.7 NS	39.14	0.545	0.572 (0.449, 0.658)	0.063 NS	8.84	0.160			
SSS	0.75	3.47 (1.51, 9.32 × 10 ⁶)	2.47 NS	866.0	3.24 × 10 ¹⁵	13.4 (0, 32.1)	-86.2***	62.02	0.866	0.041 (0, 0.156)	-0.452***	99.85	0.909			
SSSi	1	1.07 (1.03, 1.19)	0.072 NS	4.17	0.094	82.8 (62.1, 101.6)	-18.0 NS	12.01	0.205	0.422 (0.38, 0.466)	-0.078***	5.22	0.161			
GSM	0.99	0.98 (0.96, 1.00)	-0.018 NS	1.42	0.023	99.6 (87, 114.5)	0.5 NS	7.30	0.073	0.499 (0.467, 0.533)	-0.001 NS	3.34	0.033			
GSMi	1	1.05 (1.01, 1.10)	0.049*	2.49	0.059	72.3 (57.4, 84.5)	-28.1***	9.54	0.289	0.414 (0.395, 0.44)	-0.085***	2.94	0.172			
CSM	1					104.8 (76, 136.2)	5.0 NS	14.97	0.164	0.502 (0.466, 0.547)	0.002 NS	4.14	0.042			
CSMi	1					69.7 (48.4, 88.2)	-31.2**	16.92	0.333	0.413 (0.385, 0.442)	-0.088***	3.54	0.178			

Prop is the proportion of simulations whose fits were included in the analysis. Individual fits were excluded if they failed to converge or if they gave uninterpretable parameters. Fecund is the seed production per cm² basal area, u is the distance parameter and p is the scale parameter of the 2Dt distribution. Estimate is the median of the estimates from replicate simulations (the true values in all cases are fecund = 1, u = 100, p = 0.5), with 95% CI. Bias is the mean difference between the true and estimated parameter value divided by the true value. CV is the coefficient of variation of the estimated parameter value. RRMSE is the relative root mean squared error, that is, the root mean squared error divided by the true value. The asterisks indicate the statistical significance of the bias: *** P < 0.001, ** P < 0.01, * P < 0.05.

Table 3. Mean and 95% upper and lower credible intervals of model parameter estimates for seed dispersal in *Jacaranda copaia*

Distance	Fecundity				Fecundity		Dispersal		Dispersal	
	μ	σ	u	P	Mean	CI	Mean	CI	Mean	CI
Fittype	Median	Mean	Mean	CI	Mean	CI	Mean	CI	Mean	CI
SSSi	15.3	28.2	-3.180	(-2.501, -5.469)	2.745	(1.879, 4.434)	89.175	(26.746, 243.325)	0.712	(0.571, 1.124)
GSM	17.9	62.8	-2.907	(-2.875, -3.009)	2.000	(1.652, 2.405)	98.613	(28.276, 238.507)	0.588	(0.452, 0.759)
CSM	26.8	339.5	NA	NA	1.618	(1.505, 1.738)	190.3	(82.599, 403.792)	0.521	(0.367, 0.722)

SSSi refers to the summed seed shadow model with immigration, GSM is the gene shadow model, and CSM is the competing sources model. Median and mean are dispersal distances, respectively. Fecundity estimates are mean (μ) of log seed production per unit basal area (mm^2) and SD (σ), dispersal parameter estimates, u and p are the distance and shape parameters for Clark 2Dt dispersal kernel, respectively (see text).

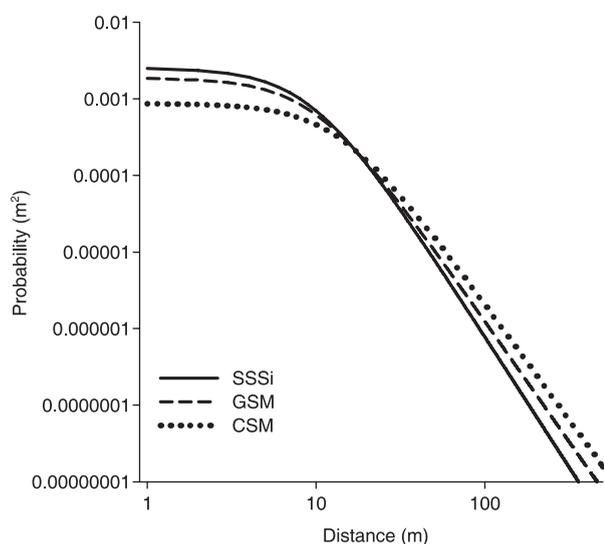


Fig. 2. Dispersal kernel estimates for the summed seed shadow model with immigration (SSSi), the gene shadow model (GSM), and competing sources model (CSM) fit to seed dispersal data from *Jacaranda copaia*.

credible intervals all overlapped (Table 3). The resulting best point estimates suggest differing pictures of dispersal potential in this species. The SSSi model produced the lowest dispersal distance estimates (median = 15.3, mean = 28.2 m), followed by the GSM (17.9, 62.8), and CSM (26.8, 339.5) (Fig. 2). The population-level fecundity parameter estimates also have overlapping credible intervals and are broadly congruent, though the best estimates of the population mean (μ) log fecundity parameter vary as SSSi > GSM, and those of the population SD (σ) log fecundity parameter vary as SSSi > GSM > CSM.

Discussion

COMPARING THE ACCURACY AND PRECISION OF THE DIFFERENT FITTING MODELS

Estimates from simulated data sets clearly demonstrated the substantial biases troubling naïve analyses of both genetic and non-genetic data, and showed how these can be overcome.

Estimation of the population dispersal distance distribution directly from the distribution of dispersal distances in a genotyped sample (OBS) leads to overestimation of dispersal distances when sampling is biased against areas with high densities of seeds. Weighting each data point inversely with the sampling effort in its source area (e.g. the proportion of seeds genotyped in a randomly placed seed trap) (OBSw) provides a simple correction for this bias, albeit the resulting estimates are still subject to considerable error depending on the locations of seed traps relative to reproductive trees.

Application of inverse modelling to seed trap data without consideration of immigrant seed rain also leads to overestimation of dispersal distances and fecundities whenever the data include immigrant seeds arriving from outside the study area. Including immigrant seed rain adds a mass conservation constraint that explicitly considers seeds that are dispersed long distances in the calculations. If the fitted seed shadow suggests significant numbers of seeds produced on the plot are dispersed off the plot, then conversely it shows that significant numbers of the seeds arriving on the plot are produced off the plot for plots in continuous forests.

Failure to consider immigrant seed rain leads to two inter-related problems. First, immigrant seeds are misattributed to trees on the plot, and thus the expected seed rain from on-plot trees must be increased (through, for example, increases in fecundity estimates) to fit the data under the assumption of no immigrant seed rain. Second, and more importantly, seed rain to distances that fall outside of the plot is unconstrained. Dispersal parameter estimates that suggest a large percentage of seeds that are going off the plot are evaluated only given what they predict for seed rain on the plot. Imagine, in a worst-case scenario, dispersal parameters that are fit when immigrant seed rain is ignored might predict 90% of seeds produced on the plot are dispersed off the plot, yet the analysis assumes that 0% of the seeds are immigrants. This leads to over-estimation of fecundity parameters (Clark *et al.* 1998) and to over-estimation of the proportion of seeds going long distances whenever there are trees outside the plot that are the sources of immigrant seeds. As we show in our simulations, including estimates of immigrant seed rain in the expected seed rain calculations (SSSi) greatly reduces the biases for the fecundity and dispersal scale parameters and reduces error.

Unexpectedly, the simulations also revealed that incorporating immigrant seed rain led to increased error for the genetic inverse modelling estimates (GSMi vs. GSM, CSMi vs. CSM). Given that genetic data allows for the identification of immigrant seeds, dispersal kernels can be better estimated when these seeds are excluded from analysis, assuming remaining seeds are correctly assigned to parents on the plot. Including immigrant seeds in the GSMi and CSMi adds more uncertainty, because dispersal distance estimates of expected total immigrant seed rain are less precise than estimates of expected seed rain from a particular parent tree. Excluding immigrant seeds also eliminates potential bias towards longer dispersal distances arising from genotyping errors that misidentify seeds as immigrants.

Also unexpectedly, all inverse modelling methods that incorporate immigrant seed rain (SSSi, GSMi and CSMi) have some small bias in the dispersal shape estimates, and in other parameters as well. These biases did not appear in simulations in which the seed sources outside the study area were numerous and evenly distributed across the landscape, suggesting that the biases are related to the additional unexplained variation in seed rain due to the unknown locations (and fecundities) of source trees outside the study area (results not shown). Clearly, it is always better to have more information on the locations and fecundities of as large a proportion of trees contributing seeds to the seed traps as possible. In the absence of such information, the better course for inverse modelling with genetic data is to drop immigrant seeds from the analyses (GSM and CSM), and the better course when using seed trap data alone is to account for them (SSSi).

Our simulation results also provided information on the relative precision of the different methods, as reflected in the confidence intervals and coefficients of variation of the estimates. These results should not be interpreted in terms of absolute precision when applied to any data set, because the precision depends on the size of the data set, the appropriateness of the fitted model, and the noisiness of the data (represented in the simulations by the negative binomial dispersion parameter). Indeed, the relative precision of the genetic and non-genetic estimates depends specifically on the proportion of seeds genotyped, and thus the relative size of the two data sets. Our simulation results are particularly relevant to the case study species we analyse, as the data sets are similar in size and dispersal scale. For data sets of these sizes, methods based on inverse modelling of genetic data have the highest precision (lowest CVs) when both dispersal parameters were examined together, with the GSM and GSMi slightly better than the CSM and CSMi for these simulated data sets. When we fit models to data sets lacking information on variation in fecundity among individuals, however, the CSM and CSMi had higher precision than the GSM and GSMi, indicating that the CSM is more robust to a poor fecundity model (results not shown), a strength also highlighted by Robledo-Arnuncio & García (2007). Across all models, precision of the dispersal shape parameter estimates was higher than that of the dispersal scale parameter estimates. GSM and GSMi estimates of fecundity were very precise, while SSSi estimates of fecundity were not.

For our simulations, the root mean squared error, which reflects both bias and error (both accuracy and precision), was lowest for the GSM and CSM models, and next best for the SSSi, GSMi, CSMi and OBSw models, and poor for the strongly biased OBS and SSS models. Again, these results should not be interpreted too broadly, as the rankings will vary with the data set.

INSIGHT GAINED INTO SEED DISPERSAL AND PRODUCTION IN *JACARANDA*

When applied to our empirical data for *Jacaranda*, all the models agree that long-distance dispersal is frequent in this species. Best estimates of the fraction of seeds going beyond 100 (1000) m vary from 3.5% (0.13%) for the SSSi to 6.5% (0.43%) for the GSM to 12.6% (1.2%) for the CSM. Considering that a mature reproductive *Jacaranda* tree produces on the order of 100 000 seeds per year, a typical tree is expected to send hundreds if not thousands of seeds beyond 1 km every year.

The dispersal parameter estimates we obtained here through intensive analyses of genetic and non-genetic data combined differ from previous estimates we obtained from these data sets using different models. Previous estimates from the genetic data set using OBS suggested larger dispersal distances, consistent with the upward bias expected under this model given the seed sampling strategy (Jones *et al.* 2005). Estimates from fitting the 2Dt with the p parameter set to a larger seed trap data set using SSS suggested a lower frequency of seeds going long distances, consistent with the bias expected given the thinner-tailed kernel being fitted (Muller-Landau *et al.* this issue). The fits presented here are free of both these biases, and should be a more accurate and precise reflection of *Jacaranda's* seed dispersal.

Overall, the differences between the SSSi, GSM, and CSM dispersal parameter estimates for *Jacaranda*, which are larger than those observed in the fits to simulated data sets, suggest that the fitted two-parameter dispersal kernel still may not fully capture the complex shape of the true dispersal kernel. Because inverse modelling estimates based on genetic data should provide better information about dispersal at distances beyond the maximum nearest tree-trap-distances, here 106 m, we expect the GSM and CSM estimates to be a better basis for drawing conclusions about long-distance dispersal in particular.

In addition to our genetic estimates, several lines of evidence suggest that the potential for LDD in *J. copaia* is high. The maximum average wind speed during the period when seeds are dispersed in *J. copaia* (August–December) is half the maximum average wind speed of the period when all other wind-dispersed tree seeds on the FDP are released (February–May (Paton 2007)). Nevertheless, *J. copaia* is one of the most effective dispersers on the FDP. It is one of only two species on the FDP that has successfully dispersed seeds to each of the 200 seed traps over a 15-year period (S. J. Wright, personal communication). This suggests that other mechanisms besides maximum wind speed, such as the vertical movement of seeds in updraft thermals, may play an important role in determining

dispersal distances. Hundreds of seeds were observed rising vertically in conditions of relatively little horizontal wind to heights > 25 m above the canopy before moving horizontally (F. A. Jones, personal observation). Moreover, *J. copaia* is one of the most common seeds caught in a network of vertical seed traps placed above the canopy (35 m) at several sites around BCI (R. Nathan *et al.*, unpublished data). If uplifting events are common, these could create complex dispersal kernels with separate distributions describing dispersal far from the plant that would not necessarily be well described by dispersal kernels with monotonic decline in seed number within increasing distance (Nathan *et al.* 2002).

Fitted models agree with empirical data in indicating considerable variation in seed production per basal area among trees. Based upon capsule counts of 188 individual *Jacaranda* individuals in this population, we estimated that the average seed numbers produced per basal area was 0.56 with a $SD = 1.12$. These estimates overlap with those obtained under the SSSi and GSM models; both population-level and individual tree estimates are in greater agreement with the GSM results (data not shown).

EVALUATING GENETIC AND NON-GENETIC ESTIMATION OF DISPERSAL

Genetic methods for collecting data on seed dispersal patterns often require significant investment of time and resources, which makes their widespread application limited. However, our results show that they can be useful for estimating dispersal distances in natural populations and can offer more precise estimates of fecundity and dispersal than non-genetic methods. Genetic models will provide the greatest additional information in cases in which there is extensive overlap in seed shadows, which reduces the precision of non-genetic analyses. An additional major advantage of high-resolution genetic data is that it makes it possible to accurately estimate rates of immigration. Identification of seeds arriving from outside the study area has revealed that long-distance dispersal is more frequent in some populations than previously recognized or estimated based upon local dispersal kernels. Finally, they can provide deeper insight on the mechanisms responsible for long-distance movement (Pairon *et al.* 2006; Jordano *et al.* 2007).

Our simulation results show that non-genetic seed shadow estimates are relatively unbiased if immigrant seed rain is explicitly considered in the model (SSSi rather than SSS), and the fitted model is a sufficiently good representation of the true dispersal kernel. However, failure to include seed immigration can result in greatly overestimated fecundity and dispersal. Clearly, the degree of such overestimation will depend on the number of immigrant seeds in the data set, which itself depends on the scale of seed dispersal relative to the scale of the study area and the density of source trees outside the study area. As demonstrated by Clark *et al.* (1998), excluding potential source trees at the edge of the study area has little effect on estimates for species with very short dispersal distances, but substantial effect for species with longer dispersal distances.

Of the techniques employed here, the SSS and SSSi are by

far the more common and can be expected to remain so in the foreseeable future. Because data for multiple species can easily be collected at the same time, relevant analytical tools are well developed and widely used, and many studies have already been completed, application of these model allows for systematic comparisons within and among plant communities across multiple years (Clark *et al.* 2004; Muller-Landau *et al.* this issue). Furthermore, if one is interested mainly in predicting expected seed rain within the study area and quantifying associated seed limitation rather than inferring dispersal distances from fitted models, the SSS and SSSi models, which effectively fit spatial variation in seed densities in the study area, provide the most direct approach. The SSSi model also provides an unbiased estimate of dispersal distances, and we strongly recommend that this model supercede the non-immigrant SSS model in all such analyses. However, even with the SSSi model, care must be taken in extrapolating results of locally derived models to larger distances.

The wide credible intervals on dispersal parameter estimates for *Jacaranda* under all the models used here provides a reminder of the difficulty of obtaining precise information on seed dispersal. Our seed trap and genotype data sets for this species are among the largest such data sets yet for any species: the area mapped for adult trees is nearly 100 ha, 300 seed traps captured over 5000 seeds with large variation in seed arrival among traps, and over 400 seeds were genotyped. Further, because *Jacaranda* is wind-dispersed, we expect its dispersal kernel to be more regular than those of the animal-dispersed species for which seed rain is much more patchy (Muller-Landau & Hardesty 2005; Russo *et al.* 2006). Yet without prior knowledge of the shape or scale parameters of the dispersal kernel, or any of the fecundity parameters, the posterior credible intervals on the parameters generally encompassed a factor of two or more. Estimates of long-distance dispersal frequency and mean dispersal distances from studies for which data sets are more limited, adult trees are more abundant, plots are smaller in scale, and/or fitted models provide an even rougher caricature of the true process are likely to suffer even greater errors. It is always possible to achieve narrower confidence intervals on fitted parameters by constraining one or more of the parameters, or fitting a simpler model – but at this time we lack a good basis for applying such a constraint, and inappropriate choice of a simplified model would lead to biased estimates (Muller-Landau *et al.* this volume).

CONCLUSIONS AND FUTURE DIRECTIONS

The integration of multiple types of data and analyses using hierarchical Bayesian models offers the best hope for further progress in understanding seed dispersal, especially long-distance dispersal (Clark 2005). This should involve not only seed trapping and genetic matching – preferably with a higher frequency of seeds genotyped in traps far from adults – but also the use of information on dispersal processes to constrain the shape of the dispersal kernel, and data on seed production of individual trees to constrain the fecundity model. For example,

in wind-dispersed species, data on tree heights combined with mechanistic models for how height affects dispersal distances could be incorporated into the fitted dispersal kernels along with information on wind speed and direction and spatial heterogeneity (Katul *et al.* 2005; Bohrer *et al.* this issue).

In order to deal with the problem of censored data in the tail due to the unknown origin of immigrant seeds, representative genetic sampling of populations distributed across a landscape of interest and the use of assignment techniques based upon more traditional measures of population genetic structure could be used to determine the most-likely source population of immigrant seeds (see review in Manel *et al.* 2005). Knowledge of the source of immigrant seeds would lessen the uncertainty of immigrant dispersal distances and allow for its integration into the tail (Clark *et al.* 2003). In the case of *Jacaranda*, genotypic information from populations across BCI could be used in combination with the assignment tests and the CSM to provide estimates of the likely sources of immigrants and provide greater resolution with regards to the source of seeds moving off plot.

Future studies should explore how to integrate missing or incomplete genotypic data into the framework presented here. Such missing data is common in the maternal tissue approach due to the degraded nature of DNA in maternal seed tissue (F. A. Jones, unpublished data). Finally, we encourage studies that integrate genetic analyses on seed movement distances with genetic analyses on seedling recruitment distances (Hardesty *et al.* 2006). Collectively, a combination of modelling, population genetic, and field techniques could shed a great deal of insight into the demographic role of LDD events within populations and its contribution to the origin and maintenance of biological diversity.

Acknowledgements

We thank J. Bullock and R. Nathan for the invitation to contribute to this special feature and for their patience. Jim Clark and an anonymous referee provided insightful and critical comments that greatly improved the manuscript. We thank S. J. Wright and O. Calderon for the use of their seed trap data, and the Minnesota Supercomputing Institute at the University of Minnesota for computing resources and user support. This work was supported by a Tupper postdoctoral fellowship from the Smithsonian Tropical Research Institute (F.A.J.), National Science Foundation grants (DEB0129874, DEB0605125, DEB043665, DEB0453445), a Packard Foundation fellowship (H.C.M.), and the University of Minnesota (H.C.M.).

References

- Bohrer, G., Katul, G.G., Nathan, R., Walko, R.L. & Avissar, R. (2008) Effects of canopy heterogeneity, seed abscission and inertia on wind-driven dispersal kernels of tree seeds. *Journal of Ecology*, **96**, 569–580.
- Brokaw, N. (1985) Gap phase regeneration in a tropical forest. *Ecology*, **66**, 683–687.
- Brooks, S. & Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434–455.
- Bullock, J.M., Shea, K. & Skarpaas, O. (2006) Measuring plant dispersal: an introduction to field methods and experimental design. *Plant Ecology*, **186**, 217–234.
- Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecology Letters*, **8**, 2–14.
- Clark, J.S. (2007) *Models for Ecological Data: An Introduction*. Princeton University Press, Princeton NJ.
- Clark, J.S., LaDeau, S. & Ibanez, I. (2004) Fecundity of trees and the colonization-competition hypothesis. *Ecological Monographs*, **74**, 415–422.
- Clark, J.S., Lewis, M., McLachlan, J.S. & HilleRisLambers, J. (2003) Estimating population spread: what can we forecast and how well? *Ecology*, **84**, 1979–1988.
- Clark, J.S., Macklin, E. & Wood, L. (1998) Stages and spatial scales of recruitment limitation in southern Appalachian forests. *Ecological Monographs*, **68**, 213–235.
- Clark, J.S., Silman, M., Kern, R., Macklin, E. & RisLambers, J.H. (1999) Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology*, **80**, 1475–1494.
- Croat, T. (1978) *The Flora of Barro Colorado Island*. Stanford University, Stanford, California.
- Dalling, J.W., Muller-Landau, H.C., Wright, S.J. & Hubbell, S.P. (2002) Role of dispersal in the recruitment limitation of neotropical pioneer species. *Journal of Ecology*, **90**, 714–727.
- Gelman, A. & Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472.
- Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (1995) *Bayesian Data Analysis*. Chapman and Hall, London.
- Godoy, J.A. & Jordano, P. (2001) Seed dispersal by animals: exact identification of source trees with endocarp DNA microsatellites. *Molecular Ecology*, **10**, 2275–2283.
- Grivet, D., Smouse, P.E. & Sork, V.L. (2005) A novel approach to an old problem: tracking dispersed seeds. *Molecular Ecology*, **14**, 3585–3595.
- Hansen, T., Brunsfeld, S., Finegan, B. & Waits, L. (2007) Conventional and genetic measures of seed dispersal for *Dipteryx panamensis* (Fabaceae) in continuous and fragmented Costa Rican rain forest. *Journal of Tropical Ecology*, **23**, 635–642.
- Hardesty, B.D., Hubbell, S.P. & Bermingham, E. (2006) Genetic evidence of frequent long-distance recruitment in a vertebrate-dispersed tree. *Ecology Letters*, **9**, 516–525.
- Hilborn, R. & Mangel, M. (1997) *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton, NJ.
- Hubbell, S.P. & Foster, R. (1983) Diversity of canopy trees in a neotropical forest and implications for conservation. *The Tropical Rain Forest: Ecology and Management* (eds S.L. Sutton, T.C. Whitmore & A.C. Chadwick), pp. 25–41. Blackwell, Oxford, England.
- Isagi, Y., Saito, D., Kawaguchi, H., Tateno, R. & Watanabe, S. (2007) Effective pollen dispersal is enhanced by the genetic structure of an *Aesculus turbinata* population. *Journal of Ecology*, **95**, 983–990.
- Jones, F.A. & Hubbell, S.P. (2003) Isolation and characterization of microsatellite loci in the tropical tree *Jacaranda copaia* (Bignoniaceae). *Molecular Ecology Notes*, **3**, 403–405.
- Jones, F.A. & Hubbell, S.P. (2006) Demographic spatial genetic structure of the Neotropical tree, *Jacaranda copaia*. *Molecular Ecology*, **15**, 3205–3217.
- Jones, F.A., Chen, J., Weng, G.-J. & Hubbell, S.P. (2005) A genetic evaluation of seed dispersal in the Neotropical tree *Jacaranda copaia* (Bignoniaceae). *The American Naturalist*, **166**, 543–555.
- Jordano, P., Garcia, C., Godoy, J.A. & Garcia-Castano, J.L. (2007) Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences USA*, **104**, 3278–3282.
- Katul, G.G., Porporato, A., Nathan, R., Siqueira, M., Soons, M.B., Poggi, D., Horn, H.S. & Levin, S.A. (2005) Mechanistic analytical models for long-distance seed dispersal by wind. *American Naturalist*, **166**, 368–381.
- Leigh, E.G., Rand, A.S. & Windsor, D.M. (eds) (1996) *The Ecology of a Tropical Forest: Seasonal Rhythms and Long-term Changes*. Smithsonian Institution Press, Washington DC.
- Manel, S., Gaggiotti, O.E. & Waples, R.S. (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, **20**, 136–142.
- Muller-Landau, H.C. & Hardesty, B.D. (2005) Seed dispersal of woody plants in tropical forests: concepts, examples, and future directions. *Biotic Interactions in the Tropics: Their Role in the Maintenance of Species Diversity* (eds D.F.R.P. Burslem, M.A. Pinard & S.E. Hartley), pp. 267–309. Cambridge University Press, Cambridge.
- Muller-Landau, H.C., Dalling, J.W., Harms, K.E., Wright, S.J., Condit, R., Hubbell, S.P. & Foster, R.B. (2004) Seed dispersal and density-dependent seed and seedling mortality in *Trichilia tuberculata* and *Miconia argentea*. *Forest Diversity and Dynamism: Findings from a Network of Large-Scale Tropical Forest Plots* (eds E.C. Losos & E.G. Leigh, Jr.), pp. 340–362. University of Chicago Press, Chicago.
- Muller-Landau, H.C., Wright, S.J., Calderón, O., Hubbell, S.P. & Foster, R.B. (2002) Assessing recruitment limitation: concepts, methods and case-studies

- from a tropical forest. *Seed Dispersal and Frugivory: Ecology, Evolution and Conservation* (eds D.J. Levey, W.R. Silva & M. Galetti), pp. 35–53. CAB International, Wallingford, Oxfordshire, UK.
- Nathan, R. & Muller-Landau, H.C. (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and Evolution*, **15**, 278–285.
- Nathan, R., Katul, G.G., Horn, H.S., Thomas, S.M., Oren, R., Avissar, R., Pacala, S.W. & Levin, S.A. (2002) Mechanisms of long-distance dispersal of seeds by wind. *Nature*, **418**, 409–413.
- Pairon, M., Jornard, M. & Jacquemart, A. (2006) Modeling seed dispersal of black cherry, an invasive tree: how can microsatellites help? *Canadian Journal of Forest Research*, **36**, 1385–1394.
- Paton, S. (2007) 2006 Meteorological and Hydrological Summary for Barro Colorado Island. Smithsonian Tropical Research Institute.
- Pounden, E., Greene, D.F., Quesada, M. & Contreras Sánchez, J.M. (2008) The effect of collisions with vegetation elements on the dispersal of winged and plumed seeds. *Journal of Ecology*, **96**, 591–598.
- Ribbens, E., Silander, J.A. & Pacala, S.W. (1994) Seedling recruitment in forests: calibrating models to predict patterns of tree seedling dispersion. *Ecology*, **75**, 1794–1806.
- Robledo-Arnuncio, J.J. & García, C. (2007) Estimation of the seed dispersal kernel from exact identification of source plants. *Molecular Ecology*, **23**, 5098–5109.
- Russo, S.E., Portnoy, S. & Augspurger, C.K. (2006) Incorporating animal behavior into seed dispersal models: implications for seed shadows. *Ecology*, **87**, 3160–3174.
- Schurr, F.M., Steinitz, O. & Nathan, R. (2008) Plant fecundity and seed dispersal in spatially heterogeneous environments: models, mechanisms and estimation. *Journal of Ecology*, **96**, 628–641.
- Tufto, J., Engen, S. & Hindar, K. (1997) Stochastic dispersal processes in plant populations. *Theoretical Population Biology*, **52**, 16–26.
- Wang, B.C., Sork, V.L., Leong, M.T. & Smith, T.B. (2007) Hunting of mammals reduces seed removal and dispersal of the Afrotropical tree *Antocaryon klaineana* (Anacardiaceae). *Biotropica*, **39**, 340–347.
- Wenny, D.G. & Levey, D.J. (1998) Directed seed dispersal by bellbirds in a tropical cloud forest. *Proceedings of the National Academy of Sciences USA*, **95**, 6204–6207.

Received 3 August 2007; accepted 14 April 2008

Handling Editor: James Bullock